# Negation detection in text using syntactic dependencies

Written by Mário Csaplár, supervised by doc. Ing. Ivan Polášek, PhD.

STU FIIT

SLOVENSKÁ TECHNICKÁ
UNIVERZITA V BRATISLAVE
FAKULTA INFORMATIKY
A INFORMAČNÝCH TECHNOLÓGIÍ

## Conclusions

In this work, we created a language model based on the syntactic structure of Slovak language. Dependency relations were the most crucial aspect for defining rules of negation detection.

A custom Slovak dataset with over 1300 manually tagged negations served as a means for evaluation.

Our dictionary approach to negation marker detection achieved F1 score of 97.45%. The dependency approach to negation scope detection achieved F1 score of 85.89%.

Negation detection module can be easily incorporated into any larger natural language processing chain.

## Introduction

### Role of negation in NLP

Negation adds higher level of semantics into automatically processed text.

„Patient _denies_ chest **pain**.“
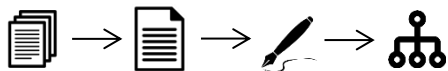„Patient _experiences_ chest **pain**.“

### Two parts of negation

1. _Negation marker_
2. **Negation scope**

„We left home _without_ **sadness**.“

## Process flow

### Preprocessing



We start with a collection of documents. First, plain text has to be extracted. Then morphological and syntactic marking is performed. At last, dependency tree is built.

### Negation detection

_Negation markers_ → **Negation scopes** → Marked documents

## Negation marker detection

### Invariant markers

Easiest to detect. Mostly used with genitive case, such as _bez_ (without), _okrem_ (apart from), _mimo_ (outside of), _namiesto_ (instead of). Also contains the word _nie_ (not).

### Inflected markers

Triggered by negative prefix _ne-_ or foreign prefixes _a-_, _anti-_, _de-_, _dys-_ and many other.

The presence of a prefix, however, is not definitively indicative, e.g. _neznámy_ (unknown) but not nežný (gentle).

Dictionary approach is to remove prefix and check for presence of prefixless word in a dictionary. Some false positives still remain.

## Negation scope detection

### Dependency parsing

A dependency tree is built according to hierarchical relationships between words on a syntactic level. The word order is not relevant. This allowed us to define negation rules in the form of tree transitions.

### Negation rules in dependency trees

Each negation category is defined with three attributes: negation marker, which direction to navigate, what word tags to look for.

An example is negation category of preposition with genitive case. Marker is the preposition, scope is the first noun or pronoun in subtrees.



„Šťastie _bez_ **závisti** nebýva.“
„Fortune never comes _without_ **jealousy**.“
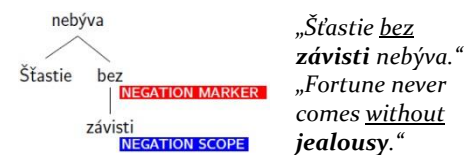
Attribute, subject's attribute (left) and no (middle) are other categories with clausal scope. Predicative negation (right) is the only one with sentence scope.
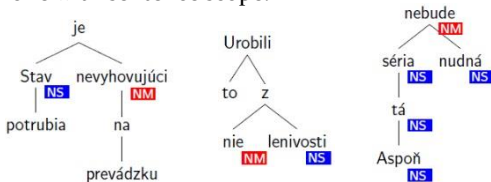


The proposed model takes into account double negatives, compound sentences and complex sentences.

## Evaluation

In this work, we created the first Slovak corpus with manually tagged negations. 4146 annotated sentences contain in total 1338 negations. Genres with various formality were included.

### Negation markers

Our dictionary approach can detect inflective markers with F1 score of 93.96%. If we also count invariant markers, total success rate is 97.45%.

### Negation scope

Across all negation categories and text genres, F1 score for negation scope detection using dependencies is at 85.89%. Experiments proved formal, technical texts provide best results.

## Practical use

The outcome is a Java library which can mark negations in Slovak and English text. We attempted to incorporate it into existing sentiment analysis and knowledge extraction systems.

## References

We contribute our success to long-term research by doc. Pavlovič, especially his work Negácia v jednoduchej vete, Slavisticky kabinet SAV, 2003.